Recognition by Association: Within- and Cross-modality Associative Priming with Faces and Voices

Sarah V Stevenage*, Sarah Hale, Yasmin Morgan & Greg J Neil

* Please send correspondence to:

Dr Sarah Stevenage,

School of Psychology; University of Southampton, Highfield, Southampton, Hampshire, SO17 1BJ, UK

Tel: 02380 592234; Fax: 02380 594597; email: svs1@soton.ac.uk

This is the accepted version of the following article: Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. British Journal of Psychology, 105(1), 1-16. doi: 10.1111/bjop.12011, which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1111/bjop.2013.105.issue-1/issuetoc

Abstract

Recent literature has raised the suggestion that voice recognition runs in parallel to face recognition. As a result, a prediction can be made that voices should prime faces and faces should prime voices. A traditional associative priming paradigm was used in two studies to explore within-modality priming and cross-modality priming. In the within-modality condition where both prime and target were faces, analysis indicated the expected associative priming effect: The familiarity decision to the second target celebrity was made more quickly if preceded by a semantically related prime celebrity, than if preceded by an unrelated prime celebrity. In the cross-modality condition, where a voice prime preceded a face target, analysis indicated no associative priming when a relatively short stimulus onset asynchrony (SOA) was used. However, when a longer SOA was used, providing time for robust recognition of the prime, significant cross-modality priming emerged. These data are explored within the context of a unified account of face and voice recognition which recognises weaker voice processing than face processing.

Recognition by Association: Within- and Cross-modality Associative Priming with Faces and Voices

Our current understanding of voice recognition has been shaped heavily by the literature on face recognition, both in terms of experimental methodologies, and in terms of the consideration of underlying processes. Recent work, however, has suggested that rather than simply reflecting isomorphic recognition systems, face and voice recognition might be more profitably viewed as parallel pathways within a single multimodal person recognition system. A priming paradigm provides a valuable method to test this prediction as it addresses the assumptions of a supramodal representation that can be accessed by each input modality (Shah, Marshall, Zafiris, Schwab, Zilles, Markowitsch & Fink, 2001). The present paper seeks to complement the existing demonstration of cross-modality identity priming through demonstration of cross-modality associative priming. In addition, however, the predictable influence of a weaker voice recognition route was explored.

A growing literature exists to support the emergent view that faces and voices sit as parallel processing pathways within a single multimodal recognition framework. This is well articulated in an overview by Belin, Bestelmeyer, Latinus & Watson (2011), and in the model of voice perception put forward by Belin Fecteau & Bedard, 2004). Through these formulations, it is clear that whilst faces and voices may activate specific areas of the brain, their representations nevertheless interact to inform identity-based decisions. This perspective is informed by three distinct literatures. First, applied behavioural studies have shown an interaction between faces and voices. For instance, Sheffert and Olson (2004) report that the capacity to learn a voice is facilitated by simultaneous presentation of the face. Similarly, several researchers note that when recognising a voice, performance may be facilitated if the face is also seen at study and at test (Legge, Grosmann & Pieper, 1984; Armstrong & McKelvie, 1996; Yarmey, 2003). Equally, more applied work highlights the capacity for one input to interfere with the processing of the other (Cook & Wilding, 1997; McAllister, Dale, Bregman, McCabe & Cotton, 1993; Stevenage, Howland & Tippelt, 2011). In all cases, these data demonstrate the capacity for face and voice processing to influence one another, as would be expected through placement as parallel pathways within the same system.

The second line of work to reveal an interaction between faces and voices rests on neuropsychological techniques. Work already exists to demonstrate voice recognition in the absence of face recognition (prosopagnosia – Shah *et al.*, 2001), and face recognition in the absence of voice recognition (phonoagnosia – Neuner & Schweinberger, 2000), and this fuelled the suggestion that any combination of inputs might occur after recognition had taken place. More recently, however, fMRI work has begun to indicate the combination of processing across sensory channels at a much earlier stage. Indeed, activation in the auditory cortex is modulated by seeing the lip movements of a speaker (Besle et al., 2009), and likewise, activation in the visual fusiform face area can be demonstrated following presentation of the human voice (von Kriegstein et al., 2005).

In a similar vein, the results of Schweinberger and colleagues are relevant. They used audiovisual integration (AVI) to explore the impact of multimodal face-voice presentations on subsequent voice recognition. Across three studies with stimuli presented in synchrony or near-synchrony, results indicated that when the voice was paired with its corresponding face a benefit was evident in terms of facilitated voice recognition. However, when the voice was paired with a noncorresponding face, a significant cost was evident in terms of voice recognition. These results have been demonstrated using both behavioural measures such as accuracy (Robertson & Schweinberger, 2010; Schweinberger, Robertson & Kaufmann, 2007) and more recently through ERP recordings (Schweinberger, Kloth & Robertson, 2011). The important point here is that, whilst the authors take these results to reflect the need for temporal contiguity or near-contiguity in the presentation of faces and voices, they are also useful in speaking to the issue of integration across modalities. Indeed, the time course of these data suggest that sensory integration may occur prior to the recognition of each input (see Kayser & Logothetis, 2007 for a review) and this, importantly, may underpin facilitation or priming effects.

The third line of work explicitly explores priming directly as a method to provide a robust and powerful test of multisensory processing. The prediction is that, if faces and voices sit within a multimodality recognition framework, the presentation of an appropriate prime stimulus should facilitate processing of a subsequent test stimulus no matter what modality those stimuli are. These effects have already been demonstrated using an *identity priming* method but are yet to be demonstrated using an *associative priming* method. Both demonstrations are important in verifying the interaction between modalities that arises from a multimodality framework. However, both demonstrations are complementary in that they rest on slightly different mechanisms and thus yield slightly different predictions.

Identity Priming

Identity priming refers to the facilitation gained when recognising a particular individual if that individual has been presented previously. Identity priming has already been demonstrated across a number of studies using within-modality tests involving (i) faces at prime and test stage (Bruce, Carson, Burton & Kelly, 1998; Bruce & Valentine, 1985; Ellis, Young & Flude, 1990), and (ii) voices at prime and test stage (Schweinberger, Herholz & Stief, 1997). However, critical to a model with parallel pathways, the literature has also demonstrated *cross-modality* priming. For instance, prior presentation of name can facilitate subsequent recognition of that person's face (Calder & Young, 1996; Burton, Kelly & Bruce, 1998), and prior presentation of a face can facilitate subsequent recognition of their name (Young, Hellawell & de Haan, 1988). More pertinently for the present study, prior presentation of a face can facilitate subsequent recognition of that person's voice (Ellis, Jones & Mosdell, 1997; Schweinberger *et al.* 1997; Stevenage, Hugill & Lewis, 2012). Hence, a parallel pathway for voices is supported.

Associative Priming

In contrast to identity priming, associative priming occurs when the presentation of one person facilitates the later processing of another person to whom they are semantically associated. This facilitation rests on the activation of some shared semantic information. It is by virtue of this shared information, that the prime celebrity provides some (back-) activation to the associated target celebrity resulting in a quicker response when that target is subsequently presented. By its nature, associative priming should be demonstrated both within-modality (the face of one person primes the face of another: Bruce & Valentine, 1986), and across-modalities (i.e., the name of one person primes the face of another; Burton, Kelly & Bruce, 1998; Schweinberger, 1996; Wiese & Schweinberger, 2008). It has yet to be demonstrated with voices.

However, given that associative priming rests on the activation of shared semantic information, the effect is presumed to be located later in the processing framework and effects are smaller and less long-lived (Burton, Bruce & Johnston, 1990). This gives rise to an important consideration because, by this stage, any relative weakness of one pathway compared to another will be magnified. Given a body of literature suggesting a weaker ~~void~~ voice recognition pathway compared to the face recognition pathway, the consequence as activation falls through propagation loss is that fewer voices may be recognised and fewer associations to semantic information can be formed and later activated. Thus, a voice prime may elicit more 'familiar only' experiences (Ellis et al., 1997; Hanley, Smith & Hadfield, 1998) and may show greater difficulty in forming and then activating associated semantic information (Barsics & Brédart, 2011; Brédart, Barsics & Hanley, 2009; Damjanovic & Hanley, 2007, 2009) than a face prime. Consequently, the prediction underlying the present paper is that if voices also represent a parallel pathway within this person-recognition model, then voices too should elicit cross-modal associative priming effects. However, these effects may be fragile, and demonstrable only under robust prime recognition conditions.

Experiment 1: Method

Design

A 2 x 2 mixed design was used in which prime modality was manipulated between-participants (within-modality, cross-modality), and prime type (related, unrelated) was manipulated within-participants. With this design, participants were presented with a prime stimulus, which was immediately followed by a target stimulus to which participants gave a speeded familiarity decision. Accuracy and speed of correct response represented the dependent variables.

Participants

A total of 41 participants completed the study in return for course credit. Ages ranged between 18 and 29 years, and all participants were familiar with all celebrity stimuli, and had normal or corrected-to-normal hearing and vision. Participants were randomly assigned to either the within-modality condition (n = 20, 16 females, mean age = 20.1 years SE = .54), or the cross-modality condition (n = 21, 11 females, mean age = 23.5 years, SE = .79).

Materials

Three sets of stimuli were used: 10 related celebrity pairs, 10 unrelated celebrity pairs, and 20 unfamiliar pairs. In all stimulus sets, the first member of the pair (prime) was always a celebrity, while the second member of the pair (target) was either a celebrity (famous) or not (unfamiliar). The primes were drawn from stage, screen and television, and were paired with the targets in a way that did not afford any obvious strategic linkages. In this way, the nature of the prime was unlikely to help the participant to predict the nature of the response to the subsequent target.

Stimuli consisted of 70 celebrities and 20 unfamiliar targets. These were combined to form 40 stimulus pairs as follows. First, 20 celebrity faces were selected from a larger set on the basis of their familiarity ratings and these were designated as 'famous' targets. Across 8 judges, these targets had a mean familiarity level of 6.36 (SD = .76, Min = 5) on a 7-point scale. Half were paired with 10 highly familiar and associated primes to form 10 related pairs, and judges' ratings confirmed an acceptable ~~levels~~ level of association (Mean = 5.85, SD = .87, Min = 4.3) on a 7 point scale. These associations designated people who were naturally seen together through their relationship (i.e., the couple Victoria and David Beckham) or through a working association (i.e., the TV presenters Ant and Dec). Thus, they represented associated rather than merely semantically (or categorically) related pairs (see Ellis, 1992). The remaining targets were paired with 10 highly familiar but unrelated primes to form 10 unrelated pairs. Identity of the targets used in related and unrelated pairs was counterbalanced across participants, necessitating an associated prime for each of the 20 targets as well as the 10 unrelated primes. This counterbalancing ensured that results could not be attributed to item effects, as each item was presented within a related pair and within an unrelated pair across participants. Associative priming would be demonstrated if each target was recognised better when preceded by the associated celebrity than when preceded by the unrelated celebrity. In both the related pair trials and the unrelated pair trials, the correct response to the target in the speeded familiarity task would be ‘famous’.

Commented [GJN1]: Should this be familiar?

The 20 unfamiliar faces served as targets in the unfamiliar pair trials. These were paired with a further 20 celebrity stimuli, selected to be highly familiar, recognisable from both face and voice, but unrelated to all other stimuli. In these unfamiliar pair trials, the correct response to the target in the speeded familiarity task would be ‘unfamiliar’.

Within-modality trials were created by presenting a *face* as prime, and as subsequent target; whilst cross-modality trials were created by presenting a *voice* as prime and a *face* as subsequent target[1].

Faces: The celebrity faces were drawn from internet sites, and depicted the celebrity in a full-frontal pose with a natural smiling expression. The unfamiliar faces were drawn from an internet modelling site, so as to be matched with the celebrity images for photographic quality, age, and general level of attractiveness. All images were edited within Corel PhotoPaint to remove all background details, and were

[1] The reverse conditions characterised by voices as primes followed by voices as targets (within-modality) and by faces as primes followed by voices as targets (cross-modality), would not have been effective in the current design as the measure of RT in the speeded familiarity task to targets would have been compromised by the time taken to present the target voice.

converted to greyscale, and matched for size based on inter-ocular distance (set to 50 pixels). Images were presented within a white 7 x 7 cm square such that the face itself measured approximately 3.3 x 4.7 cm.

Voices: The celebrity voices were extracted from YouTube interview clips and thus represented segments of free, rather than scripted, speech. Clips were edited to be 3 seconds in length and, in line with the recommendations of van Lancker, Kreiman and Emmorey (1985) and Schweinberger *et al*. (1997), care was taken to ensure that speech content did not reveal the identity of the speaker (as confirmed by the lack of ability of judges to identify each speaker from a written transcript).

The experiment was conducted on a Toshiba laptop running Windows Vista. Faces were viewed on a 13" colour monitor with a screen resolution of 1280 x 800 pixels, and at a viewing distance of approximately 60 cm. Voices were presented via the computer speakers within a quiet environment, ensuring good audibility. Stimulus presentation and data collection were controlled using SuperLab v2.0.

Procedure

Participants were tested individually within a quiet cubicle, and the task was introduced to them as an exploration of how much one person can influence the identification of another. Online instructions prepared participants for the sequential presentation of pairs of stimuli. Participants in the within-modality condition saw a prime face followed by a test face, while participants in the cross-modality condition heard a prime voice followed by a test face. In both conditions, the participant was asked to attend to, but not respond to, the prime stimulus. However, for the test face, they were asked to indicate, as quickly but as accurately as possible, whether it was *familiar* or *unfamiliar*. Responses were made by pressing labelled keys (M for 'familiar', and Z for 'unfamiliar') on a standard computer keyboard, and 20 practice trials ensure that the participant had adequately mapped each response to each keypress.

The 40 experimental trials were presented in two blocks of 20 separated by a self-paced break. In both the within-modality and cross-modality conditions, a central fixation cross was first presented for 200 ms to orient the participant. The prime was then presented for 3 seconds, followed by a visual mask for 40 ms, and then the target face which remained visible until response. In this way, the stimulus onset asynchrony (SOA) was held constant at 3040 ms in within- and cross-modality conditions. The capacity to recognise the prime at this exposure duration was noted, and the speed and accuracy of the familiarity decision to the target represented the dependent variables. The order of related, unrelated, and unfamiliar trials was randomised within each block and the entire testing sequence lasted no more than 15 minutes.

Following completion of the experimental trials, participants completed a post-experimental questionnaire. All participants were asked to view each celebrity face with the aim of providing a name or other unique identifying information. Additionally, they were asked to rate the celebrity faces for degree of association to their celebrity pair. Ratings were made on a 7 point scale where 1 indicated low association, and 7 indicated high association.

Cross-modality participants were also asked to identify the celebrity voices that they had heard as prime stimuli within their trials. However, this was completed after each experimental trial in order to minimise the extent that rumination about a prior voice might affect the response speed to a subsequent trial. Thus, cross-modality participants were asked whether the voice they had heard within each trial was familiar to them or not. If they indicated it to be familiar, participants were encouraged to provide a name or other unique identifying information as above.

These data were important to the experimental design in that if a prime celebrity was not known, or if no perceived association between prime and target existed for the participant, priming to a related target celebrity could not be expected (see Brunas-Wagstaff, Young & Ellis, 1992 for the importance of spontaneous recognition in priming). Thus, trials in which either the prime or the target stimuli were not known were removed from subsequent analyses on a case-by-case basis.

Results and Discussion

Accuracy of response on the speeded familiarity task was noted as a way of screening out participants who performed at chance levels or below when responding to the target face. On this basis, the data from 5 participants in the within-modality condition and 3 participants in the cross-modality condition were excluded without replacement. In addition, priming could not be anticipated if the prime had not been identified. With a 3 second exposure duration to the prime, an imbalance in identification rates was clear: all face primes were identified but performance with voice primes was notably worse at just 29% identification rate. Data were omitted from participants on each trial where there was a failure to meet a 50% threshold identification of voice primes either by name or unique semantic information. This resulted in the omission of 5 further participants and 54.7 % of data across the remaining participants in this condition.

Accuracy of Response

Accuracy of performance on the speeded familiarity task is summarised in Table 1. A 2 x 2 mixed Analysis of Variance (ANOVA) was conducted using prime type (related, unrelated) as the within-participants variable, and modality (within, cross) as the between-participants variable. This revealed neither a significant effect of prime type ($F_{(1, 26)} = 1.31$, *ns*) nor modality ($F_{(1, 26)} < 1$, *ns*). No interaction emerged to qualify these results ($F_{(1, 26)}$ = 3.89, *ns*,), thus accuracy was uniform across all conditions.

(Please insert Table 1 about here)

Response Speed

Associative priming was also investigated through the more usual analysis of response speed, in which priming would be demonstrated through a speed-up of performance to a target following a related prime than following an unrelated prime. The data are summarised in Table 1, using median RT as the dependent variable to minimise the influence of outliers in the data. As above, a 2 x 2 mixed ANOVA showed no effect of prime type overall ($F_{(1, 26)} < 1$, *ns*). However a main effect of modality did emerge ($F_{(1, 26)}$ = 5.05. $p > .05$, $\eta^2 = .16$) with faster responding in the within-modality condition than the cross-modality condition overall. This was qualified by a significant interaction between prime type and modality ($F_{(1, 26)}$ = 8.16, $p < .01$, $\eta^2 = .24$), and Bonferroni-correct post-hoc explorations confirmed a significant priming effect in the within-modality condition ($t_{(14)} = 2.49$, $p_{(1\text{-tailed})} = .013$) but not in the cross-modality condition ($t_{(12)} = 1.64$, *ns*)[2].

The results here suggest the presence of associative priming in the within-modality (face-face) condition, and this was to be expected from the literature. The present results, however, did not reveal any associative priming in the cross-modality (voice-face) condition either when examining accuracy or more traditionally speed of familiarity judgement. This was a surprise given the predictions from the literature.

Formatted: Subscript

---

[2] Although item effects were well controlled within the experimental design, the results reported here were replicated in a 'by-items' analysis. The ANOVA revealed a main effect of modality ($F_{(1, 19)}$ = 69.87, $p < .001$, partial $\eta^2 = .79$) with longer RTs in cross-modality than within-modality conditions. In addition, there was a significant main effect of prime type ($F_{(1, 19)} = 9.53$, $p < .01$, partial $\eta^2 = .33$) showing the expected associative priming effect. The absence of a significant interaction ($F_{(1, 19)}$) < 1, *ns*) masked the fact that there was priming under within modality conditions ($t_{(19)} = 2.61$, $p_{1\text{-tailed}} = .01$) but not under cross-modality conditions ($t_{(19)} = 1.37$, *ns*).

However, critical to the demonstration of associative priming is the capacity to identify the prime stimulus. Without this, no benefit could be expected.

In this regard, the findings of Schweinberger, Herholtz and Sommer (1997) are of value. In an elegant investigation, Schweinberger *et al.* explored the effect of exposure duration on recognition decisions (famous/not famous). Their data suggested that an increase of exposure duration for a voice beyond 1 second had little additional effect when the task was to recognise the speaker as ‘famous or not’. However, they noted that even the maximal exposure duration of 2 seconds yielded an identification performance (by name or unique semantic identifier) of only 4% across 40 celebrity voices.

The use here of a 3 second exposure duration for the face and the voice prime provided a greater opportunity to recognise the voice primes compared to Schweinberger et al.’s exposure duration. Moreover, it was sensible to provide a test of priming effects where the SOA between prime and target was held constant. However, whilst 3 seconds was ample for the identification of the face prime, it was clearly still not adequate for the identification of the voice prime. With this in mind, Experiment 1 may not have provided a suitable test of cross-modal priming and Experiment 2 is presented to address this issue.

Experiment 2: Method

Recognition rates from voice clips have been reported across a number of studies now. Whilst some studies show acceptable levels of recognition from short clips, these generally test recognition from a closed list of speakers in which the listener is asked to identify a speech clip from a given number of individuals. For example, Abberton and Fourcin (1978), ~~round~~ found a near 100% recognition rate across 11 participants recognising 5 matched familiar voices merely saying “Hello, how are you?”.

In contrast, when an open list of speakers is used, the length of voice clip required for acceptable recognition is somewhat higher. For instance, Ladefoged and Ladefoged (1980) explored the recognition of 29 personally familiar speakers interspersed with 24 once-heard or unfamiliar speakers. Their results suggested 31% recognition from a single word ("hello"), 66% recognition from a single sentence, and 83% recognition from a 30 second voice clip. Thus, whilst recognition could be achieved, the level of recognition is sensitive to the length of clip used.

In order to maximise the potential for recognition of the voice prime, the exposure duration to the voice clip was extended from 3 seconds to 8 seconds. This change to the methodology rests on pilot work, and on the results of a series of published studies which indicates a recognition rate from 8 second clips of between 58% and 63% across celebrity voices (see Hanley & Damjanovic, 2009; Hanley & Turner, 2000) and once-heard voices (Stevenage, Clarke & McNeill, 2012; Stevenage Howland & Tippelt, 2011). As such, Experiment 2 provides a fairer test of cross-modal associative priming from voice to face.

Design

Cross-modality priming was explored under a longer SOA through varying the prime voice (related, unrelated) that preceded a target face. Participants gave a speeded familiarity decision to the target face, and accuracy and speed of response represented the dependent variables.

Participants

A total of 24 participants (19 females) took part in the present study. Ages ranged between 20 and 24 years (mean age = 20.96 years, SE = .15), and no participants had taken part in the previous experiment. All were familiar with all celebrity stimuli, and had normal or corrected-to-normal hearing and vision.

Materials

The materials were identical to those used in Experiment 1.

Procedure

All procedural details were identical to those used in Experiment 1 except for the change to the exposure duration for the prime voice. This was now increased to provide an 8 second clip of uninterrupted natural speech rather than the 3 second clip used previously. This change was made to encourage a higher identification rate of prime stimuli in the cross-modality condition. All other details were unchanged.

Results and Discussion

As in Experiment 1, accuracy of response on the speeded familiarity task was used as a way of screening out participants who performed at chance levels or below. On this basis, data from 4 participants were omitted from further analysis leaving data from 20 participants (2 males) remaining. In addition, data were omitted pertaining to any trials in which the prime was not recognised. Given the use of an 8 s voice clip, this resulted in an identification rate of 57.75%, which is comparable with the identification levels reported with an 8 s clip by Hanley & Damjanovic (2009; Expt 1: 64%, Expt 2: 63%) and Hanley & Turner (2000; 57%). Consequently the data from 1 further participant, and 40% of the data across remaining participants were omitted from the following analysis.

Accuracy and Speed of Response

As in Experiment 1, associative priming was investigated through analysis of accuracy and then speed of response (median RT) to the target following either a related prime or an unrelated prime. The data are summarised in Table 2 and are explored by means of two repeated-measures *t*-tests. This revealed no priming effect when accuracy was considered ($t_{(18)} < 1$, *ns*), but a clear and significant effect of priming when the more usual measure of response speed was considered ($t_{(18)} = 2.24$, $p_{(1\text{-tailed})} < .025$). This was confirmed in a 'by-items' analysis[3] ($t_{(17)} = 1.98$, $p_{(1\text{-tailed})} = .032$), suggesting a clear and significant benefit in a speeded familiarity judgement to a famous face when an associated person's voice had been presented previously. This is the first demonstration of cross-modality associative priming from voices to faces, and its demonstration here rested on the use of a paradigm that enabled sufficient spontaneous recognition of the primes.

(Please insert Table 2 about here)

General Discussion

The present paper focussed on associative priming effects and sought convergent evidence for the suggestion that voices and faces represented parallel pathways within a single multimodal person recognition system. Whilst a body of work exists to support this prediction, the relative weakness of the voice recognition pathway compared to the face recognition pathway had the potential to jeopardise a demonstration of associative priming because it had the capacity to jeopardise the recognition of the voice, and the access of semantic information associated with it. Consequently, the prediction was tested here with an expectation of a fragile effect.

Commented [GJN2]: This was the only part I had to read twice. Perhaps something like "Whilst a body of work exists to support this prediction, the voice recognition pathway may not have been strong enough to support associative priming, either because of poor voice recognition or poor access to relevant semantic information." It might even be better to split it into two sentences.

Within this theoretical context, the present results provide a clear demonstration of associative priming in the within-modality case (face prime: face target). This was to be expected and confirmed the findings from several studies within the literature (Bruce et al., 1998; Bruce & Valentine, 1985; Ellis et al., 1990; Schweinberger et al., 1997). In contrast, a significant effect of associative priming

[3] Two items were omitted from the 'by-items' analysis due to their identification as outliers in terms of median speed of response. This left 18 items.

only emerged in the cross-modality case (voice prime: face target) when the exposure duration to the prime, and hence the SOA between prime and target, was sufficiently long to enable a reasonable level of prime recognition and access of associated semantic information-. This pattern of results could not be accounted for by item-effects, as the design enabled all target celebrities to be rotated across both related and unrelated prime conditions. Rather than interpreting these data to indicate priming only when voices were given an advantage over faces through a long prime presentation, we interpret this long prime presentation as necessary to enable prime identification to an acceptable limit. In this regard, the longer prime exposure used in Experiment 2 was a response to the relative weakness in voice processing compared to face processing at the shorter 3s exposure of Experiment 1. These results thus demonstrate cross-modality associative priming, as predicted by a single multimodality framework for person recognition. However, they also demonstrate again the relative weakness of voices compared to faces as a means of identification.

The IAC Account

Several explanations may be offered to account for the present data. Perhaps the most widely cited explanation for priming effects is provided by Burton, Bruce and Johnston's (1990) IAC model. Based on Bruce and Young's (1986) earlier Information Processing model of familiar face recognition, the IAC model proposes that person recognition proceeds through activation of units at various layers of a connectionist network. In this context, face recognition is achieved when an input face activates a set of feature input units, and then a face recognition unit (FRU) which represents that face. In the IAC framework, activation at this level translates to the subjective experience of having seen a face before. When activation is successfully propagated to the next unit - the person identity node (PIN) - then a sense of familiarity is gained. The PIN is assumed to represent a modality-free representation, receiving input from multiple modalities (face, name, voice, gait, etc) and from this representation, the retrieval of semantic information and a name can be achieved.

Priming sits well within this explanatory framework but the account differs according to whether the long-lived identity priming effect, or the shorter-lived associative priming effect is being described. Identity priming, for example, is explained through appealing to a link-strengthening argument in which prior exposure to a target fortifies the links between one layer of units and the next, facilitating their activation at subsequent exposure. Thus the presentation of a prime face activates the FTUs, FRU and PIN associated with that face, and strengthens all links between them so that subsequent activation is faster. Cross-modality identity priming effects can similarly emerge through the fact that when the PIN associated with a prime voice becomes active, activation can spread back to facilitate processing along a parallel modality pathway associated with the test face. In both cases, facilitation can be observed.

The explanation for associative priming is somewhat different and recognises the fact that associative priming occurs beyond the level of the PIN, is weaker, and is less long-lived. Rather than appealing to a notion of link strengthening, associative priming is explained with reference to the notion of spreading activation. More specifically, the presentation of a prime results in the activation of a PIN and then of semantic information associated with that PIN. With some semantic information being shared, as is the case when people are highly associated, activation can be propagated back to trigger a low level of activation in the PIN of an associated person. It is this residual activation that makes it easier to recognise the associated target person when they are subsequently presented. In this regard, it should not matter *how* the PIN of the prime is activated. The only factor of importance is *that* the PIN of the prime is activated. Both within-modality and cross-modality associative priming are predicted, and accounted for, in this way.

What is appealing about this explanation is the capacity for the IAC model to recognise and reflect the weakness of the voice pathway relative to the face pathway. With voices being harder to identify than faces, it is possible that the PIN is activated less from a voice input than from a face input, resulting in weaker identification, and weaker subsequent semantic information retrieval underpinning associative priming. However, Hanley and Damjanovic (2009) provide data which challenge this account inasmuch as semantic retrieval remains worse from voices than from faces even when recognition is matched through blurring, and familiarity is matched through a well controlled repeated-measures design. In this instance, a reformulated IAC account may still be able to account for differential effects across faces and voices if multisensory integration is presumed to occur at the semantic level rather than at the level of the familiarity decision.

The Strategic Processing Account

An alternative explanation for priming effects, especially at relatively long SOAs as used here, is articulated by Wiese (2011). Wiese reviews a very appealing literature which describes an alternate and more strategic account of priming. Under a long SOA for instance, expectancy-based processing (Becker, 1980) is possible in which participants may show facilitated performance to a target because they have the capacity to strategically predict its presentation once they have learned that primes and targets are sometimes related.

Whilst a short SOA may minimise this strategic approach, strategic effects are nevertheless possible at short SOAs through a second process – retrospective semantic matching (Neely, Keefe & Ross, 1989). Here, participants at the test stage may retrospectively test whether the target they see is related to the preceding prime. If so, the test response is clarified – in the current case, the response should be 'famous'. Hence a fast 'famous' response may occur because of the retrospective detection of the association between prime and target rather than being purely driven by the perceptual processing of the target itself. These strategic influences do not negate the demonstration of priming here. However, they offer an alternative account which is more in line with the episodic memory accounts of Jacoby (1983, see also Hintzman, 1986; Logan, 1988; Roediger, 1990; *cf* Bruce, Carson, Burton & Kelly, 1998). The attraction of the structural account, and the strategic accounts, however, is that they are both better able to explain the impact of relative pathway strength on the resulting priming effect.

A Neuropsychological Approach

This discussion is not complete without some reflection on recent studies which reveal priming effects through ~~EPR~~ ERP methods. These studies have the advantage of separating out the perceptual priming (revealed through N100) from the potential involvement of strategic processing (possible at N250 and beyond). Moreover, these studies remove the need to equalise face and voice prime recognition through artificial means such as face degradation through blurring, or extended voice exposure. In this regard, the relative impact of differential pathway strength on perceptual priming may be examined through the use of sensitive neuropsychological measures.

With this in mind, the work of Föcker, Hölig, Best and Röder (2011) presents an exciting avenue because of its parallels to the present work. In an innovative series of studies, Föcker *et al.* explored the crossmodal interaction between faces and voices

Commented [GJN3]: Familiar?

Commented [GJN4]: Familiar?

within a self-priming paradigm. Using behavioural measures, they revealed self-priming effects under both within-modality (voice-voice) and cross-modality (face-voice) conditions. More interestingly, using neuropsychological measures, they revealed self-priming, (through a lower ERP waveform to person-congruent trials than person-incongruent trials) at both an N100 and at a later N270-530 component. Importantly, all effects were significant in the cross-modality case but were rather more fragile and perceptually rather than semantically based in the within-modality (voice-voice) case at an early stage. These data suggest the multisensory integration at very early stages of processing, whilst being sensitive enough to reveal differences across face prime and voice prime conditions. This points to the utility of ERP measures as a way of exploring the subtle differences across modalities without degrading or otherwise altering the stimulus presentation. In this vein, the results of an ERP study using voices at prime and faces at test in a semantic priming study would be of value as a way of providing a neuropsychological correlate to the present behavioural demonstration.

Conclusion

The present paper has provided a demonstration of both within- and cross-modality associative priming between voices and faces. These data converge with those from a repetition priming paradigm, and together confirm the utility of a single multimodal model of person recognition in which faces and voices represent separate parallel input pathways. Moreover, the current results speak to the issue of the relative fragility of the voice recognition pathway compared to the face recognition pathway inasmuch as prime identification was harder to achieve from a voice prime than a face prime. Both structural accounts and strategic accounts of priming may be advanced to account for the present results, and neuropsychological correlates for these findings are eagerly anticipated as a way of potentially distinguishing between these accounts.

References

Abberton, E., & Fourcin, A.J. (1978). Intonation and speaker identification. *Language and Speech, 21*, 305-318.

Armstrong, H.A., & McKelvie, S.J. (1996). Effect of Face Context on Recognition memory for Voices. *The Journal of General Psychology, 123(3)*, 259-270.

Barsics, C., & Brédart, S. (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition, 20(2)*, 303-308.

Becker, C.A. (1980). Semantic context effects in visual word recognition – An analysis of semantic strategies. *Memory and Cognition, 8(6)*, 493-512.

Belin, P., Bestelmeyer, P.E.G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*, 711-725.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Science, 8*, 129-135.

Besle, J., Betrand, O., & Giard, M.H. (2009). Electrophysiological (EEG, sEEG MEG) evidence for multiple audiovisual interaction in the human auditory cortex. *Hearing Reserach, 258*, 143-151.

Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology, 21*, 1013-1021.

Brédart, S., Valentine, T., Calder, A., & Gassi, L. (1995). An interactive activation model of face naming. *Quarterly Journal of Experimental Psychology, 48A(2)*, 466-486.

Bruce, V., Carson, D., Burton, A.M., & Kelly, S.W. (1998). Prime time advertisements: Repetition priming from faces seen on subject recruitment posters. *Memory and Cognition, 26(3)*, 502-515.

Bruce, V., & Valentine, T.R. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76*, 373-383.

Bruce, V., & Valentine, T. (1986). Semantic priming of familiar faces. *Quarterly Journal of Experimental Psychology, 38A(1),* 125 – 150.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Brunas-Wagstaff, J., Young, A.W., & Ellis, A.W. (1992). Repetition priming follows spontaneous but not prompted recognition of familiar faces. *Quarterly Journal of Experimental Psychology, 44(3),* 423-454.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81*, 361-380.

Burton, A.M., Kelly, S.W., & Bruce, V. (1998). Cross-domain repetition priming in person recognition. *The Quarterly Journal of Experimental Psychology, 51(A),* 515-529.

Calder, A.J., & Young, A.W. (1996). Self priming: A short-term benefit of repetition, *Quarterly Journal of Experimental Psychology 49A*, 845–861.

Cook, S., & Wilding, J. (1997). Earwitness Testimony 2: Voices, Faces and Context. *Applied Cognitive Psychology, 11,* 527-541.

Damjanovic, L., & Hanley, J.R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory and Cognition, 35,* 1205-1210.

Ellis, A.W. (1992). Cognitive mechanisms of face processing. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences, 335(1273)*, 113-119.

Ellis, A.W., Young, A.W., & Flude, B.M. (1990). Repetition priming and face recognition: Priming occurs within the system that response to the identity of the face. *Quarterly Journal of Experimental Psychology, 42(A),* 495-512.

Ellis, H.D., Jones, D.M., & Mosdell, N. (1997). Intra- and Inter-Modal Repetition Priming of Familiar Faces and Voices. *British Journal of Psychology, 88*, 143-156.

Föcker, J., Hölig, C., Best, A., & Röder, B. (2011). Crossmodal interaction of facial and vocal personal identity information: An event-related potential study. *Brain Research, 1385,* 229-245.

Hanley, J.R., & Damjanovic, L. (2009). It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory, 17*, 830-839.

Hanley, J.R., Smith, S.T., & Hadfield, J. (1998). I recognise you but can't place you. An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology, 51A(1),* 179-195.

Hanley, J.R., & Turner, J.M. (2000). Why are familiar-only experiences more frequency for voices than for faces? *Quarterly Journal of Experimental Psychology, 53A(4),* 1105-1116.

Hintzman, D.L. (1986). Schema abstraction in a multiple trace memory model. *Psychological Review, 93*, 411-428.

Jacoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 21-38.

Keyser, C. & Logothetis, N.K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structures and Function, 211,* 121-132.

Ladefoged, P., & Ladefoged, J. (1980). The ability of listeners to identify voices.

UCLA Working Papers in Phonetics, 49*, 43-51*.

Legge, G.E., Grossman, C., & Pieper, C.M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning Memory and Cognition, 10*, 298-303.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychology Review, 95*, 492-527.

McAllister, H.A., Dale, R.H.I., Bregman, N.J., McCabe, A., & Cotton, C.R. (1993). When Eyewitnesses are also Earwitnesses: Effects on Visual and Voice Identifications. *Basic and Applied Social Psychology, 14(2),* 161-170.

Neely, J.H., Keefe, D.E., & Ross, K.L. (1989). Semantic priming in the lexical decision task – Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15(6),* 1003-1019.

Neuner, F., & Schweinberger, S.R. (2000). Neoropsychological impairments in the recognition of faces, voices and personal names. *Brain Cognition, 44*, 342-366.

Robertson, D.M.C., & Schweinberger, S.R. (2010). The role of audiovisual asynchrony in person recognition. *Quarterly Journal of Experimental Psychology, 61(1),* 23-30.

Roediger, H.L. (1990). Implicit memory: retention without remembering. *American Psychologist, 45,* 1043-1056.

Schweinberger, S. R. (1996). How Gorbachev primed Yeltsin: Analyses of associative priming in person recognition by means of reactions times and event-related brain potentials. *Journal of Experimental Psychology, 22(6),* 1383 – 1407.

Schweinberger, S.R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Lanugage and Hearing Research, 40*, 453-463.

Schweinberger, S.R., Herholz, A., & Stief, V. (1997). Auditory Long-term Memory: Repetition Priming of Voice Recognition. *Quarterly Journal of Experimental Psychology, 50A(3)*, 498-517.

Schweinberger, S.R., Kloth, N., & Robertson, D.M.C. (2011). Hearing facial identities: Brain correlates of face-voice integration in person identification. *Cortex, 47(9)*, 1026-1037.

Schweinberger, S.R., Robertson, D., & Kaufmann, J.M. (2007). Hearing facial identities. *Quarterly Journal of Experimental Psychology, 60*, 1416-1456.

Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001). The neural correlates of person familiarity: A functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804-815.

Sheffert, S.M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception and Psychophysics, 66(2)*, 352-362.

Stevenage, S.V., Howland, A., & Tippelt, A. (2011). Interference in Eyewitness and Earwitness Recognition. *Applied Cognitive Psychology, 25(1)*, 112-118.

Stevenage, S.V., Hugill, A.R., & Lewis, H.G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, 24(4), 409-419.

Van Lancker, D., & Kreiman, J. & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13*, 19-38.

Von Kriegstein, K., Kleinschmidt, A., & iraud, A.L. (2005). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex, 16*, 1314-1322.

Wiese, H. (2011). The structure of semantic person memory: Evidence from semantic priming in person recognition. *British Journal of Psychology, 102*, 899-914.

Wiese, H., & Schweinberger, S.R. (2008). Event-related potentials indicate different processes to mediate categorical and associative priming in person recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1246-1263.

Yarmey, A.D. (2003). Earwitness identification over the telephone and in field settings. *International Journal of Speech, Language and the Law, 10(1)*, 62-74.

Young, A.W., Hellawell, D., & deHaan, E.H.F. (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40A(3)*, 561-580.

Appendix 1:

Associated celebrity pairs used in Experiments 1 and 2.

Victoria and David Beckham

Ant and Dec

David Mitchell and Robert Webb

Kirstie Alsopp and Phil Spencer

Rob Brydon and James Cordon

Fearne Cotton and Holly Willoughby

Sharon and Ozzy Osbourne

Bruce Forsythe and Tess Daly

Gordon Ramsay and Jamie Oliver

Justin Lee Collins and Alan Carr

Katie Price and Peter Andre

Dawn French and Jennifer Saunders

Trinny and Susannah

Kim and Aggie

Jeremy Clarkson and Richard Hammond

Richard and Judy

Phil Schofield and Fern Britton

Simon Cowell and Louis Walsh

Patsy Palmer and Sid Owen (‘Bianca’ and ‘Ricky’ from UK TV Soap ‘Eastenders’)

Matt Lucas and David Walliams

Table 1: Accuracy and median RT for correct decisions (and standard error) following related or unrelated primes and under within- and cross-modality conditions in Experiment 1 when prime exposure was 3 seconds.

	Related Accuracy	Unrelated Accuracy	Related RT (msecs)	Unrelated RT (msecs)
Face Prime	87.3 (3.2)	78.7 (4.6)	764 (105)	895 (120)
Voice Prime	86.2 (3.4)	88.5 (4.9)	1245 (113)	1133 (129)

Table 2: Accuracy and median RT for correct decisions (and standard error) following related or unrelated primes and under cross-modality conditions in Experiment 2 when prime exposure was 8 seconds.

	Related Accuracy	Unrelated Accuracy	Related RT (msecs)	Unrelated RT (msecs)
Voice Prime	89.5 (1.8)	90.5 (2.4)	1079 (130)	1221 (101)